# Critical analysis of Big Data challenges on similar sequential Algorithm based data Search

Yuhua Peng[1,*], Wenli Xu[1], and Yan Xiong[1]

[1] *Wuchang University of Technology Artificial Intelligence school, Wuhan ,Hubei ,430223 , China*
[*] *Corresponding author. E-mail: yuhuapeng@yandex.com*

*Received: Apr. 5, 2021, Accepted: Apr. 25, 2021*

For improvement of cycle method of force of china, there is need to organize and recuperate the prevalence of force source within the network, during this examination, as indicated by the non-intermittent and occasional qualities of the consistent public list of force greatness, an effect predominance consistent state list appraisal and conjecture framework established on turbulent plan hypothesis and littlest squares arrangement vector component in huge information foundation is planned. Initially, tumultuous framework hypothesis is used to remake the stage planetary of the authentic information of traditional force greatness consistent state files, and to make another information data space covering attractors. At that time, the LSSVM is employed to arrange the examples in high-dimensional space, and also the (PSO) calculation is used together to urge the simplest list assessment and expectation framework model. Simultaneously, the framework is applied to the important observing of the electrical energy treatment limit of a circulation network during a specific spot. The regular consistent state record of force quality is employed to assess and screen, and therefore the normal relative mistake is under 7%. Clearly, the end result is superior to the customary back spread (BP) neural organization forecast strategy, which demonstrates that the control class consistent state list assessment framework established on turbulent framework model and statistical procedure uphold course machine underneath enormous information may be broadly utilized.

**KEYWORDS:** Chaotic system theory; Power quality steady state index; Large data; Provision vector mechanism; Big data

http://dx.doi.org/10.6180/jase.202108_24(4).0021

## 1. Introduction

In computer science, sorting and scanning are two basic operations. Sorting involves grouping the details in the sequence in which it rises or decreases. In a set of objects, searching implies locating a position or locating an aspect of a given object. Many data constructions are used to store knowledge, but simple data structures used for sorting and searching process are arrays, connected lists and tree. Any type of arithmetical data, script, series, and charm data is the search feature. A variety of search systems, such as consecutive search, dualistic search, tree exploration and hashing, have been industrialized. Each search procedure depends on the specific challenge, data stuff, and difficulty of the procedure. This research paper offers a brief introduction to the search algorithm, describes which form of search algorithm is used for which form of query, and contrasts various types of search algorithms in a various significant parameter such as complexity of time, complexity of room, related key, no similarities, etc. The element category originates from a file or chart. Each part is a call to best. The search method refers to a folder or bench and seeks a place for a particular document. A key is connected with each document and this key is split into various documents. If the key is stored at the beginning of the log, the internal key is considered this sort of key [1]. In some instances, the key is placed on a different table with a reference to the database, so this sort of important is considered an outside key.

Each of the files or tables has two main sets. This key is a purported main key and this collection includes an interior and outside key. The first collection determines specific details [2]. The second collection determines none of the special knowledge that this key is a subordinate key. We look for an component with a position that we need to set first afterward that.By the quick growth of China's budget and the incessant growth of science and technology, great and significant fluctuations have occupied place in the characteristics of power grids, load composition and control techniques of power systems in the 21st century. Both distribution and power generation are developing towards automation and distributed energy structure [3].

## 2. Methodology

For the evaluation, monitoring and prediction of the main indicators of control quality, the change trend of control quality at the corresponding monitoring points can be obtained in advance, which causes the corresponding process of power grid and the attention of managers to potential control quality problems, so as to provide correct decision-making for the management and protection of power quality problems. At present, in China's control quality forecasting, monitoring and evaluation system, data collection and collation are carried out by day, week, month, season and year. Quantitative statistics is often used as index data with probability values approaching 96% [4]. According to previous data statistics, it is found that the trend change of control excellence steady-state index has the characteristics of quasi-periodic and non-periodic. In chaos theory, this trend can be restored by spatial reconstruction. For this reason, a prediction system model founded on chaotic system theory and LSSVM is designed. The design process is shown in figure 1. In the figure above, the process of control quality steady-state index evaluation and prediction system mainly includes the following steps [5].

Firstly, the facts of control quality steady-state index evaluation and prediction system are pre-processed, and the abnormal data are judged, corrected, and normalized. According to Laida theorem [6], the abnormal data information is identified and the steady-state index arrangement x= xi/i=1,2,n in control quality assessment is obtained.

The corresponding mean value $\bar{x}$ and residual error VI are obtained.

$$v_i = x_i - \bar{x}(i = 1, 2, \cdots, n) \quad (1)$$

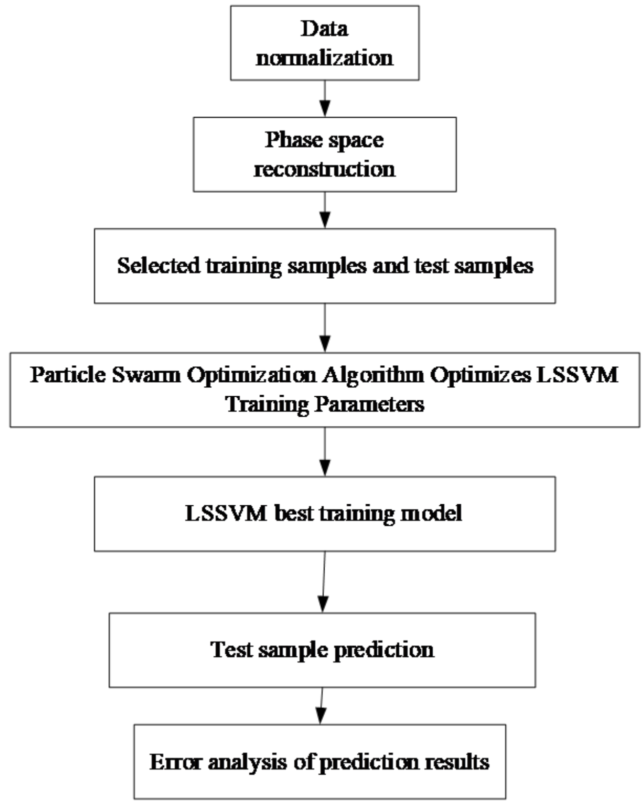


**Fig. 1.** Flowchart of forecasting model of control quality steady state indices.

According to Bessel expression, the corresponding standard error is calculated. If the detection value $xk(1 \le k \le n)$ satisfies [7]:

$$|v_k| = |x_k - \bar{x}| > 3\sigma \quad (2)$$

Then, xk is judged as an abnormal value, and the average value of the values in the two monitoring points at the adjacent time points is identified and corrected [7]. Secondly, the optimal mosaic frequency m and the optimal delay time point $\tau$ are obtained by the corresponding improved C-C method [8], and the phase interplanetary reconstruction of the steady state index of the electric energy treatment quantity is performed. At the same time, the largest Lyapunov exponent is intended by the minor data technique [9], and the chaotic model of the sequence is entered. Thirdly, according to the phase space trajectory expression theory in chaos theory [10], the training sample and the test sample are taken. Fourthly, the LSSVM is used to test and train the newly extracted samples in the high-dimensional region, and the PSO algorithm is used to find the optimized normalized numerical parameters C and numerical parameters fifthly, the trained LSSVM system [11] is used to evaluate the steady state index in the control quality assessment, and the corresponding error analysis operation is performed. The main links of the steady state index evaluation and prediction system in the control quality assessment mentioned above is phase region reconstruction and LSSVM and parameter optimization, which will be elaborated on the following.

### 2.1. LSSVM parameter optimization based on PSO

In the procedure of exercise and monitoring the LSSVM, the values of normalized numerical parameter C and numerical kernel parameter $\delta$ are particularly important. In order to optimize the evaluation and prediction system, the PSO algorithm is used to catch the optimal numerical limit C and the numerical kernel limit $\delta$. In PSO, random particles update the corresponding velocity and position relationship by searching discrete extreme worth Pbest and group thrilling worth gbest. In this way, after repeated updates, the global optimal answer of the collection is finally found. The updating formula of particle velocity and position information is as surveys.

In the formula, vt and xt respectively refer to the speed and coordinates of the particle's t-th update; w refers to inertial weights; r1 and r2 refer to random values in the region [ 1]; c1 and c2 refer to learning factors [12]. The detailed steps of optimizing the corresponding numerical parameters by using PSO algorithm [13] are as follows: Firstly, the numerical parameters are initialized. The information of particle size, update times, learning factors and particle accident location and speed are set. Secondly, the normalized numerical parameter C and the numerical kernel parameter $\delta$ are considered as a set of random particles. Each group of particles corresponds to aLSSVM model. In the current model, xtrain and ytrain are trained. The predicted and actual values are compared and the particle adaptation values in this coordinate are obtained. Thirdly, in the relative motion of particles, the adaptive values in different coordinates will be obtained. Each movement of the particle is then compared in size.

The relatively small adaptive values are corresponded to the corresponding coordinate positions and are taken as the current optimal coordinate positions of the particle. Fourthly, by comparing the fitness values between the particles in the same time period, the fitness values with smaller values are corresponded to the position of a particle, and then are taken as the current optimal coordinate position information in the population. At the same time, the particles are updated by formula 15 and formula 16. Fifthly, whether the all-out number of updates has been reached is checked. If it meets the corresponding conditions, it is necessary to finish the calculation and output the corresponding numerical results, or return to the second step for recalculation [14].

### 2.2. Case study and contrast test

Through the design of the voltage quality steady-state index evaluation and prediction system mentioned above, using chaos theory, and combining the optimized LSSVM algorithm and improved PSO algorithm, a smart grid voltage quality steady-state index evaluation system based on large data analysis is obtained. In order to further prove that the system can be better applied, the voltage quality steady-state index evaluation and prediction system designed in this study is applied to a substation monitoring point for test experiments. At the same time, it is compared with BP neural network method to verify its optimization degree.

In order to count the accuracy of examination and prediction, relative error Ere, average relative error Emre and root mean square error Ermse are introduced to get the following formula [23]:

$$E_{re} = \frac{F(i) - L(i)}{L(i)} \cdot 100\% \tag{3}$$

$$E_{\mathrm{rmse}} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{F(i)} - \mathrm{L(i)})^2} \tag{4}$$

In the formula, F(i) refers to the foretold value and L(i) refers to the real worth. The three formulas are used to further verify whether the BP neural network method and the optimization system involved in this study have smaller and more accurate relative error maximum, average relative error and root mean square error.

## 3. Results and discussion

In the steady-state index of control quality evaluation, the data information monitored on the day is taken as the daily index value, and the large data composed by the control excellence evaluation and prediction checking system of a certain power network is taken as the experimental sample. The index values of voltage deviation, distortion rate of total harmonics, unbalance of three-phase voltage and long-term flicker are evaluated and predicted. Taking the deviation value of voltage as a sample, the evaluation and prediction procedure of regulator quality index is described concretely. Figure 2 reflects the time series of voltage deviation daily indicators of a 10kV substation monitoring point in 2016-2018. Considering that the net structure, load classification and volume capacity of the predictive monitoring point have not altered much in a certain retro of time, the detected data can be taken as corresponding experimental samples.

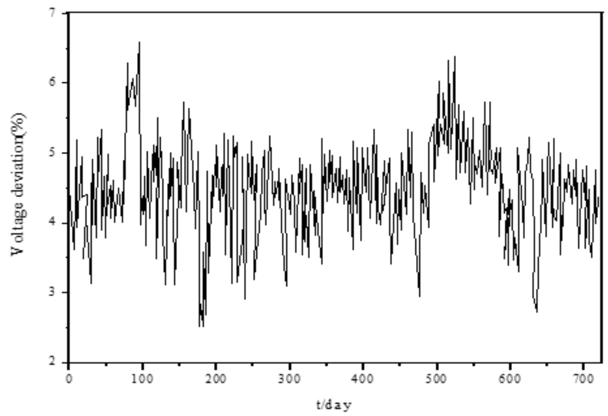


**Fig. 2.** Raw voltage deviation with the time series.

By using the improved C-C method mentioned above, the phase-area redistribution parameters of the sequence of voltage deviation time points are calculated, and the primary partial least of is taken as the optimal postponement $^{\tau}$d. The periodic opinion of is regarded as the optimal embedded window $^{\tau}$ through the formula $^{\tau}$ w=(md-1)$^{\tau}$ d, the best embedding dimension md can be obtained. The corresponding results are shown in figure 3, and $^{\tau}$ d=12.1, $^{\tau}$ w=96.2, and md=9.02 are obtained.

According to the obtained td and md, the maximum Lyapunov promoter of the time series of voltage deviation is designed by means of the small data information technique. If the end shows a positive number, the time series has certain chaotic characteristics. Figure 4 shows the Lyapunov exponential curve of the time series of voltage deviations' represents the steps of discrete time evolution, and y (k) represents the logarithmic average of distance. It can be seen that within the K interval [0,200], y (k) is approximately a straight line, and its slope is the highest Lyapunov index $^{\lambda}$. It can be calculated $^{\lambda}$ = 0.043, which indicates that the sequence of voltage deviation time points at this point has chaotic characteristics.

md = 9.02, and td = 12.1 are substituted into formula 3. For the reconstruction of phase region of voltage deviation time series, training samples (xtrain,ytrain) are adopted .LSSVM is used to train its parameters in high dimensional space.

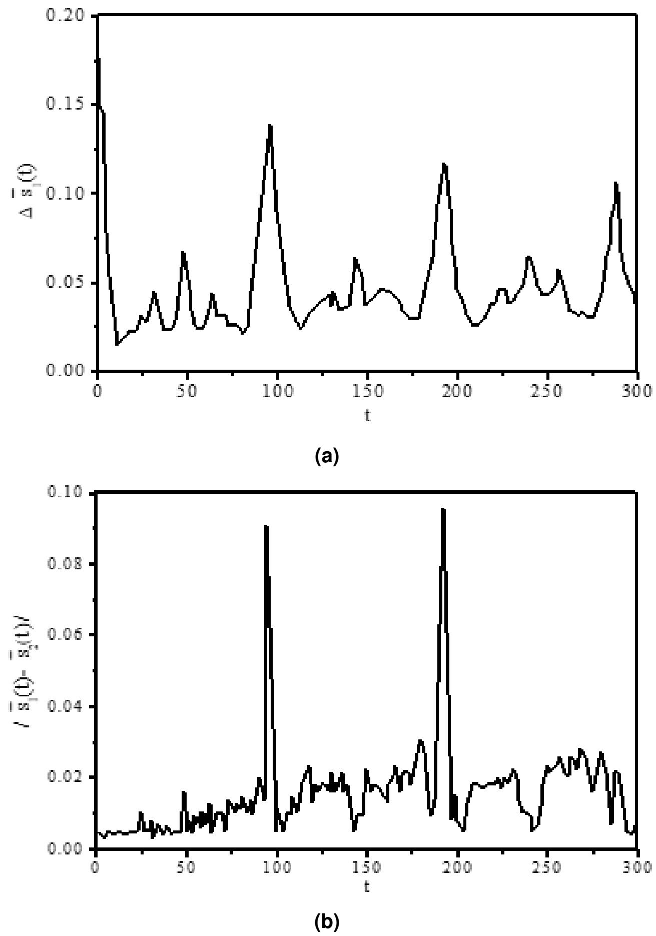


**Fig. 3.** Reconstruction parameters based on improvedC-C method.

According to PSO, the LSSVM model is optimized to obtain population number N = 30, learning factor c1 = c2 = 1.45, all-out update number Tmax = 502, and inertia weight W = 0.92. Search range C$\epsilon$ [0,502] and $\delta\epsilon$[0,50] are set. At the same time, the LSSVM toolbox is used to make some simulation comparisons with a certain web search method.

### 3.1. Comparisons with web search method

In figure 5, PSO and web search are used to enhance the strictures of LSSVM model, and the fitness is the root mean square error of the predicted and real standards in the corresponding training samples.

Thus, compared with the grid search technique, PSO has faster convergence speed on the adaptive curve, and the corresponding values are obtained: C = 73.127, $\delta$ = 0.742.

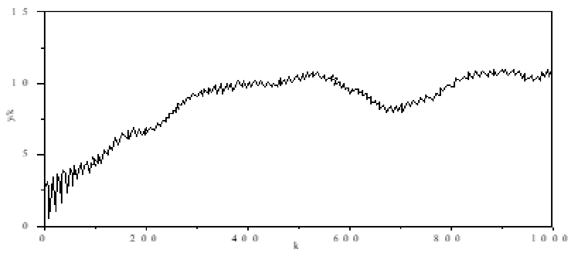

**Fig. 4.** voltage deviation time series.

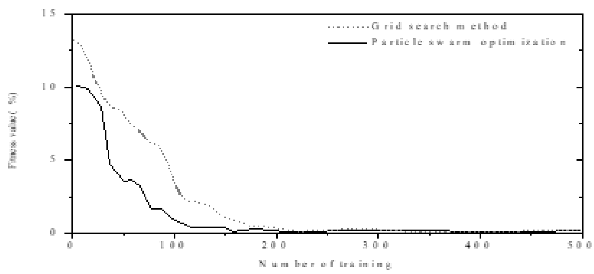


**Fig. 5.** Fitness curves based on PSO and grid search method.

### 3.2. Comparison with BP neural network

The C=73.127 and the $\delta$=0.742 obtained above are substituted into the LSSVM model. According to the corresponding chaotic theory, the reciprocal value of Lyapunov exponent $\lambda$, Tm=$1/\lambda$ refers to the upper limit of predictive monitoring time in the chaotic theory system, that is, the longest predictive time. The maximum Lyapunov exponent $\lambda$ = 0.043 for the sequence of voltage deviation time points, thus Tm = 21.23d is obtained. At the same time, on the basis of guaranteeing the prediction accuracy, the daily voltage deviation of the monitoring point in December 2018 is evaluated and predicted. In the process of forecasting, the forecasting value of each step is retained, and then it is used as the reference value for the next step. In demand to explore the rationality and practicability of the proposed assessment and prediction system model, the system and BP neural network method are compared and tested. The results are shown in figure 6.

From the figure, it can be seen that the foretold results founded on chaotic theory system and LSSVM are similar and closer to the actual values. It can be known that the evaluation system designed in this study is an evaluation and prediction system founded on BP neural network technique. Secondly, in command to further test the correctness of the system, the maximum relative error, average relative error and mean square error of the system prediction are associated with the BP neural net technique. From table 1, the steady-state index evaluation system designed in this study has advantages in the approximation of relative error, average relative error and root mean square, and this system is more accurate.

## 4. Discussion

The steady state index in control quality assessment reflects the state of power system under stable operation. After the design of the steady-state index evaluation and prediction system in control quality assessment, the chaos theory and LSSVM and PSO algorithm are used to construct the system. C-C method and small data quantization are used to enter the sequence operation. Through the corresponding empirical application analysis, compared with the actual error prediction value of network search method and BP neural network method, it is concluded that the steady-state index evaluation system in power quality evaluation in this study has great advantages.

## 5. Conclusion

Based on the related of large data, the steady-state index evaluation system for power quality evaluation of smart grid is designed and applied. Chaos theory system, LSSVM and PSO are used to construct the system. It is concluded that the system designed in this study has earlier convergence speed and training correctness than the web search technique. At the same time, it is concluded that voltage deviation, entire harmonic misrepresentation rate, three-phase voltage disturb and long-term glimmer have chaotic characteristics among the factors affecting the change of steady-state index in power quality assessment. In addition, the improved C-C technique is introduced. By scheming the optimal mosaic measurement and the optimal system postponement with local or partial minima and periodic correspondence points, the relevant features of chaotic attractors can be well mapped. At the same time, by comparing the designed system algorithm with BP neural network algorithm, it is concluded that the prediction results of local chaotic system and LSSVM model are more accurate and closer to the actual values than those of BP neural network. In addition, it is superior to BP neural net in terms of maximum relative error, root mean square error and average relative error. The average comparative error of steady-state index evaluation and prediction in control quality evaluation is less than 7%. The design of this system will

**Table 1.** Forecasting results of voltage deviation.

| Time | Actual value/% | BP neural network method | | This paper designs the system model | |
|---|---|---|---|---|---|
| | | Predictive value/% | Ere/% | Predictive value/% | Ere/% |
| 12-01 | 4.612 | 5.121 | 10.431 | 4.321 | -6.872 |
| 12-02 | 5.087 | 4.832 | -5.021 | 5.197 | 3.176 |
| 12-03 | 4.656 | 4.312 | -8.132 | 4.490 | -3.765 |
| 12-04 | 4.233 | 4.197 | -0.637 | 4.401 | 1.568 |
| 12-05 | 4.087 | 4,265 | 3.781 | 4.166 | 1.436 |
| 12-06 | 4.176 | 4.323 | 0.541 | 4.137 | -0.745 |
| 12-07 | 4.301 | 4.398 | 1.276 | 4.097 | -2.043 |
| 12-08 | 4.825 | 5,109 | -8.605 | 4.478 | -6.436 |
| 12-09 | 5.266 | 5.231 | -3.834 | 5.204 | -2.143 |
| 12-10 | 4.377 | 4.456 | 19.12 | 4.231 | -4.452 |
| 12-11 | 4.126 | 4.564 | 9.23 | 4.126 | 0 |
| 12-12 | 4.254 | 4.387 | 3.677 | 4.178 | -2.798 |
| 12-13 | 4.209 | 4.367 | 5.124 | 4.561 | 0.439 |
| 12-14 | 4.298 | 4.278 | 1.198 | 4.143 | -1.107 |
| 12-15 | 4.785 | 4.432 | -7.561 | 5.234 | 6.231 |
| Emre | - | 5.912% | | 2.834% | |
| Ermse | - | 0.324% | | 0.163% | |

bring new technology reference for the application of smart grid system based on large data background. At the same time, more and more consideration has been remunerated to the investigation of early warning system of steady-state indicators in control quality assessment. It is hoped that experts, and scholars will make joint efforts to reduce voltage and improve the quality of power supply.

## Acknowledgement

Supported by Establishment of School-level Top-quality Online Courses in 2018, WuChang Teaching Words of Technology and Teaching [2018]31, Entry name: Java Language Programming; Production-University Cooperation and Education Project of Higher Education Department of Ministry of EducationGao's letter[2018]4, Item number: 201702159003, Discussion on the Optimization and Reform of Java Course; Production-University Cooperation and Education Project of Higher Education Department of Ministry of Education, Gao's letter[2018]47, Item number:201801193096, Construction of Practice Base for Integrating Industry and Education with Computer Professional Training; Production-University Cooperation and Education Project of Higher Education Department of Ministry of EducationGao's letter[2018]47, Item

## References

[1] P O Asagba and E O Osaghae. Is Binary Search Technique Faster Than Linear Search Technique ? *Scientia Africana*, 9(2):83–92, 2010. ISSN 1118-1931.

[2] Jose Juan Dominguez Veiga, Martin O'reilly, Darragh Whelan, Brian Caulfield, and Tomas E. Ward. Feature-free activity classification of inertial sensor data with machine vision techniques: Method, development, and evaluation. *JMIR mHealth and uHealth*, 5(8), 2017. ISSN 22915222.

[3] R. Arghandeh, A. Von Meier, L. Mehrmanesh, and Et Al. On the definition of cyber-physical resilience in power systems. *Renewable and Sustainable Energy Reviews*, 58: 1060–1069, 2016.

[4] Zhihan Lv, Houbing Song, Pablo Basanta-Val, Anthony Steed, and Minho Jo. Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. *IEEE Transactions on Industrial Informatics*, 13(4):1891–1899, 2017. ISSN 15513203.

[5] Kun Wang, Jun Yu, Yan Yu, Yirou Qian, Deze Zeng, Song Guo, Yong Xiang, and Jinsong Wu. A survey on energy internet: Architecture, approach, and emerging technologies. *IEEE Systems Journal*, 12(3):2403–2416, 2018. ISSN 19379234.

[6] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob,

Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. The rise of "big data" on cloud computing: Review and open research issues, 2015. ISSN 03064379.

[7] Alexandra Von Meier, Emma Stewart, Alex McEachern, Michael Andersen, and Laura Mehrmanesh. Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications. *IEEE Transactions on Smart Grid*, 8 (6):2926–2936, 2017. ISSN 19493053.

[8] Ilhami Colak, Seref Sagiroglu, Gianluca Fulli, Mehmet Yesilbudak, and Catalin Felix Covrig. A survey on the critical issues in smart grid technologies, 2016. ISSN 18790690.

[9] Abhishek Kumar, Bikash Sah, Arvind R Singh, Yan Deng, Xiangning He, Praveen Kumar, and R C Bansal. A review of multi criteria decision making (MCDM) towards sustainable renewable energy development, 2017. ISSN 18790690.

[10] Yousef Al Horr, Mohammed Arif, Amit Kaushik, Ahmed Mazroei, Martha Katafygiotou, and Esam Elsarrag. Occupant productivity and office indoor environment quality: A review of the literature, 2016. ISSN 03601323.

[11] Shan Lin, Fei Miao, Jingbin Zhang, Gang Zhou, Lin Gu, Tian He, John A Stankovic, Sang Son, and George J Pappas. ATPC: Adaptive transmission power control for wireless sensor networks. *ACM Transactions on Sensor Networks*, 12(1):31, mar 2016. ISSN 15504867.

[12] M. Sohail Talha and H. Hajji. Analysis of research on amazon AWS cloud computing seller data security. *International Journal of Research in Engineering and Innovation*, 4 (3):131–136, 2020.

[13] Muhammad Talha - Sana Azeem - Sohail - Javed -Rabia Tariq. Mediating effects of reflexivity of top management team between team processes and decision performance. *Azerbaijan Journal of Educational Studies*, 1(1): 105–119, 2020.

[14] Osamah Ibrahim Khalaf, F Ajesh, A. A. Hamad, Gia Nhu Nguyen, and Dac Nhuong Le. Efficient Dual-Cooperative Bait Detection Scheme for Collaborative Attackers on Mobile Ad-hoc Networks. *IEEE Access*, 2020. ISSN 21693536.